

Computing Machinery I

Assignment 1

8% of your final score
Due Sept 25th @ 11:59PM MST

Lead TA

The lead TA for this assignment is: Abdelghani Guerbas (abdelghani.guerbas@ucalgary.ca)

Objective

The objective of this assignment is to practice basic C concepts, including 2D arrays, random numbers, and sorting.

Skills needed for this assignment

- Ability to work with 2D arrays in C
- Ability to use a random number generator
- Ability to employ a simple sorting algorithm
- Ability to use command-line arguments
- Ability to write and read text files in C
- Ability to define functions in C
- Ability to pass parameters by value and by reference

Note

Some of your code, if well designed, can be re-used for your project Part 1.

Overview

1. Your program will emulate a search engine. It reads a table of integers from a file **or** randomly creates the table and displays it to the user. The rows of this table represent text documents, and the columns represent words that may appear in these text documents. Each cell with coordinates $[i,j]$ in the table contains the number of occurrences of word j in document i .
2. Your program returns a list of the top n documents containing word j . The user specifies both n and j .

Details

Display to the user an $M \times N$ table of random/given positive integers between 0 and 9. The user provides M and N integers between 5 and 20 and optionally the file name containing an existing table, if applicable, using command line arguments, such as (if occurrences are randomly generated):

```
./assign1 5 6
0 1 0 2 3 1
2 5 1 0 0 0
0 7 8 0 2 7
0 0 0 5 4 0
5 8 0 0 0 9
```

In this example, there are 5 documents and 6 possible words. Document 1 contains 2 occurrences of word 0, 5 of word 1, and 1 of word 2, and no occurrences for words 3 to 5.

If occurrences are read from a file named *occurrences.txt*, then the name of the file is also specified:

```
./assign1 5 6 occurrences.txt
9 1 0 0 0 0
1 3 5 8 5 3
7 3 9 0 0 0
0 0 0 5 4 5
0 0 1 6 3 1
```

The user then enters the index of the word s/he is searching for and the number of the top documents containing the word, as follows:

```
Enter the index of the word you are searching for: 5
How many top documents you want to retrieve? 2
```

Your program will return the indices of the two documents where the frequency of the searched word (with index 5 in the example above) is the highest and the second highest among all documents.

Frequency of word j in document i is: $\text{occurrences}[i,j] / \text{size of document } i$

The size of document i (row i) is the sum of occurrences of all of its words (sum of all columns in row i).

Use any sorting algorithm. The user can quit the program or choose to search again. A log file must be created showing: the initial table, user input and the search result before exiting the program.

You do not need to worry about ties: they can be sorted in any order

Modularity

Your code must be divided into functions as appropriate. At a minimum, you must define the following functions (we are not showing all necessary arguments):

- `initialize(*table)`
- `randomNum(m,n)`; m and n are the lower and upper bounds for the random number. You can use the C library function `rand()`.
- `display(*table)`
- `topRelevantDocs(*table, n)`
- `logToFile()`

Submission

- **Note:** The lead TA may provide further submission instructions
- Name your program *assign1.c*
- Create a script file and call it *assign1.script*.
- Name your log file *assign1.log*
- Submit a *README* file providing extra instructions or information for your TA, such as the sorting algorithm you are using
- Submit your work to the appropriate dropbox on D2L.

Late Submission Policy

Late submissions will be penalized as follows:

-12.5% for each late day or portion of a day for the first two days

-25% for each additional day or portion of a day after the first two days

Hence, no submissions will be accepted after 5 days (including weekend days) of the announced deadline.

Academic Misconduct

This assignment is to be done by individual students: your final submission must be your own original work. Teamwork is not allowed. Any similarities between submissions will be further investigated for academic misconduct. While you are encouraged to discuss the assignment with your colleagues, this must be limited to conceptual and design decisions. Code sharing by any means is prohibited, including *looking* at someone else's paper or screen. The submission of compiler generated assembly code is absolutely prohibited. Any re-used code of excess of 5 lines in C and 10 lines in assembly (10 assembly language instructions) must be cited and have its source acknowledged. Failure to credit the source will also result in a misconduct investigation.

D2L Marks

Marks posted on D2L are subject to change (up or down).

Computing Machinery I

Assignment 1 Rubric

Student: _____

Item	Max Points	Points	Comments
Code compiles	5		
Code runs	5		
Correct results	15		
Log file functionality	10		
User interface (input validation, implementing all features)	15		
Random numbers	10		
Modularity	15		
Command-line arguments	5		
Passing parameters by reference	5		
Script file	5		
Code readability (formatting documentation)	10		
Total Points	100		